

README how to best use data released from the TRY database

- a) Generalities, data structure and recommendations for data extraction and import
- b) Column headers
- c) List of standardized metadata
- d) Recommendations in the case of publication

Please note: categorical traits for many species (plant growth form, woodiness, leaf type, etc.) and information about the measurement locations (soil, climate, biome-type) is publicly available at the TRY website: <http://www.try-db.org/TryWeb/Data.php>

a) Generalities and data structure

The data are provided as zipped text files (.txt) with 'latin1 swedish ci' encoding (the default setting of MySQL installations) prepared on a Linux 64-bit operating system. Columns are tab-delimited. The first row contains the header. The decimal point is given as a point (e.g. 0.123) not as a comma.

1) You receive only the trait records, which are public or for which you have the permissions from data contributors.

2) A trait may consist of different sub-groups, which have been called "Data", given as DataName and DataID.

3) You receive the original values of traits and covariates (Column: OrigValueStr). The units of some trait data and the geographical references have been standardized. In these cases you additionally receive the standardized values (Column: StdValue and Unit_1_UnitName (unit of standardized values)). The standardized values of traits and geographical references (latitude and longitude) are checked for general consistency.

4) Each row in the data table contains an entry in the database (identified by ObsDataID). This entry is categorized either as trait – in this case it has a TraitID - or as covariate – in this case it has no TraitID.

5) The "ObservationID" relates database entries (measurements of traits and covariates) on the same entity to observations (e.g. two traits measured on the same leaf). **Different rows in the data table with the same ObservationID are directly related to each other.**

6) You receive the requested trait records and all covariates related to the respective Observation. The covariates you receive may therefore not always make sense in the context of your trait data.

7) The data you receive may contain data from plants grown in natural environment and from plants grown under experimental conditions. If plants have grown under experimental conditions this is reported as a covariate entry (DataID 327, Column StdValueStr), if known. The covariate is related to the trait entry via the ObservationID. Same for mature and juvenile plants (DataID 413), and for healthy and non-healthy plants (DataID 1961).

8) The data may contain duplicates, e.g. if the same data have been contributed to TRY by different contributors. If we have identified an entry as duplicate you will find the ID of the original entry in the column OrigObsDataID.

9) A general consistency check for the standardized values of traits (Column: StdValue) has been applied. For standardized trait values we provide the largest distance of the trait record

from either species-mean, genus-mean, family mean or the mean of all data in terms of respective standard deviation ("z-score" Column: ErrorRisk). For example, an error risk value of 3 indicates that this trait record has a maximum distance of 3 standard deviations from at least one of the mean values mentioned above. Error risk > 4 may indicate likely problems in the data (similar to Kattge et al. 2011 Global Change Biology, supplementary material). The z-scores for all levels are available at: <https://www.try-db.org/TryWeb/Data.php#25>

10) Geographic coordinates are standardized to longitude and latitude values in decimal format (Column: StdValue) and checked against ESA CCI Land Cover Map of Global Water Bodies (<https://www.esa-landcover-cci.org/?q=node/162>).

11) Data extraction and import: You should be able to extract the zip file in any operating system. You can read the data file (.txt) with a text editor. It should also seamlessly open in Excel. However, in some cases the file may contain too many rows to be accessible via Excel.

To import the data into R, we recommend the 'fread' function, which also works well for large files (>100000 traits). In case of large files first install a 64-bit version of R. Then download the 'data.table' package from an R repository and make it available via 'require' and finally import the data:

```
require(data.table)
TRYdata <- fread("file.txt", header = T, sep = "\t", dec = ".", quote = "\"", data.table = T)
```

'TRYdata': example name of the R data frame into which the data will be imported; 'file.txt': path and name of the file to be imported; 'header = T': the file has a header; 'sep = "\t"': columns are tab delimited, 'dec = "."': decimals are given by a point; 'quote = "\""': text within a column is not characterized by an additional sign.

After import into R you can check the data:

dim(TRYdata): returns the number of rows and columns (number of columns should be 28)
ls(TRYdata): should return the names of headers in alphabetical order
head(TRYdata): should return the header and first rows
tail(TRYdata): should return the header and last rows

2019-Feb-12

Lucas Damian Gorné (gorneld@gmail.com) reported a bug in the R data.table package, which may cause problems: <https://github.com/Rdatatable/data.table/issues/3369>

Please do not hesitate to contact us in case you have questions with respect to the data, if you have problems importing the data, or if you find inconsistencies:

jkattge@bgc-jena.mpg.de; boenisch@bgc-jena.mpg.de

b) Column headers

Column	Comment
1. LastName	Surname of data contributor
2. FirstName	First name of data contributor
3. DatasetID	Unique identifier of contributed dataset
4. Dataset	Name of contributed dataset
5. SpeciesName	Original name of species
6. AccSpeciesID	Unique identifier of consolidated species name

7. AccSpeciesName	Consolidated species name
8. ObservationID	Unique identifier for each observation
9. ObsDataID	Unique identifier for each record
10. TraitID	Unique identifier for traits (only if the record is a trait)
11. TraitName	Name of trait (only if the record is a trait)
12. DataID	Unique identifier for each sub-trait or context information
13. DataName	Name of sub-trait or context information
14. OriginalName	Original Name of sub-trait or context information
15. OrigValueStr	Original value as text string
16. OrigUnitStr	Original unit as text string
17. ValueKindName	Value kind (single measurement, mean, median, etc.)
18. OrigUncertaintyStr	Original uncertainty as text string
19. UncertaintyName	Kind of uncertainty (standard deviation, standard error,...)
20. Replicates	Count of replicates
21. StdValue	Standardized value: available for standardized traits
22. StdUnit	Standard unit: available for standardized traits
23. RelUncertaintyPercent	Relative uncertainty in %
24. OrigObsDataID	Unique identifier for duplicate entries
25. ErrorRisk	Indication for outliers: distance to mean in standard deviations
26. Reference	Reference to be cited if trait record is used in analysis
27. Comment	Explanation for the OriginalName in the contributed dataset

Sometimes R may show a column 28, which should be empty. This column is then an artefact due to different software (MySQL >> R).

c) Standardized Metadata

Latitude:	DataID 59
Longitude:	DataID 60
Altitude:	DataID 61
Sampling or measurement date standardized:	DataID 6601
Exposition:	DataID 327
Plant developmental status / plant age / maturity / plant life stage:	DataID 413
Health status of plants (vitality):	DataID 1961
Primary reference:	DataID 113 and other

‘Primary reference’ is the reference to the original publication if data were contributed to TRY via an aggregated database (e.g. GlopNet). The number in column StdValueStr refers to the ID in the primary reference file available at: <https://www.try-db.org/TryWeb/Data.php#23>
Primary references are currently standardized up to TRY version 4.

d) Recommendations in case of publication

The intellectual property guidelines of TRY provide the following suggestions:

Any product that involves data under the Creative Commons Attribution License ‘CC BY’ must provide appropriate credit. Following the ‘Joint Declaration of Data Citation Principles’ this involves citation of the data sources, and these data citations should ‘facilitate giving scholarly credit and normative and legal attribution to all contributors to the data’ (<https://doi.org/10.25490/a97f-egykh>).

Any product that involves trait data received via the TRY Database therefore needs to cite the references of the datasets as contributed to TRY and the standard reference of the TRY Database.

In the context of peer-reviewed publications these citations should be accountable by indexing services like Google Scholar or the Science Citation Index (SCI).

If data received during their embargo period have been used in analyses, the respective custodians need to be invited to contribute to the development of manuscript/s, if they had indicated this at the time of data request.

In case of publications that rely on a significant amount of public plant trait data received via TRY, or if a research could not be completed without a specific dataset or would be notably compromised, it is strongly recommended to consider inviting the respective data custodian(s) and the TRY database management for collaboration and authorship.

Sending invitations for co-authorship at early stages of manuscript preparation is strongly recommended, so valid suggestions by co-authors can be accommodated. Co-authorship should be offered to those who respond to this invitation and subsequent updates within a timely fashion with a significant contribution.

The standard reference of the TRY Database is currently:

Kattge, J., S. Díaz, S. Lavorel, I. C. Prentice, P. Leadley, G. Bönsch, E. Garnier, M. Westoby, P. B. Reich, I. J. Wright, J. H. C. Cornelissen, C. Violle, S. P. Harrison, P. M. v. Bodegom, M. Reichstein, B. J. Enquist, N. A. Soudzilovskaia, D. D. Ackerly, M. Anand, O. Atkin, M. Bahn, T. R. Baker, D. Baldocchi, R. Bekker, C. Blanco, B. Blonder, W. J. Bond, R. Bradstock, D. E. Bunker, F. Casanoves, J. Cavender-Bares, J. Q. Chambers, F. S. Chapin, J. Chave, D. Coomes, W. K. Cornwell, J. M. Craine, B. H. Dobrin, L. Duarte, W. Durka, J. Elser, G. Esser, M. Estiarte, W. F. Fagan, J. Fang, F. Fernández-Méndez, A. Fidelis, B. Finegan, O. Flores, H. Ford, D. Frank, G. T. Freschet, N. M. Fyllas, R. V. Gallagher, W. A. Green, A. G. Gutierrez, T. Hickler, S. Higgins, J. G. Hodgson, A. Jalili, S. Jansen, C. Joly, A. J. Kerkhoff, D. Kirkup, K. Kitajima, M. Kleyer, S. Klotz, J. M. H. Knops, K. Kramer, I. Kühn, H. Kurokawa, D. Laughlin, T. D. Lee, M. Leishman, F. Lens, T. Lenz, S. L. Lewis, J. Lloyd, J. Llusià, F. Louault, S. Ma, M. D. Mahecha, P. Manning, T. Massad, B. Medlyn, J. Messier, A. T. Moles, S. C. Müller, K. Nadrowski, S. Naeem, Ü. Niinemets, S. Nöllert, A. Nüske, R. Ogaya, J. Oleksyn, V. G. Onipchenko, Y. Onoda, J. Ordoñez, G. Overbeck, W. A. Ozinga, S. Patiño, S. Paula, J. G. Pausas, J. Peñuelas, O. L. Phillips, V. Pillar, H. Poorter, L. Poorter, P. Poschlod, A. Prinzing, R. Proulx, A. Rammig, S. Reinsch, B. Reu, L. Sack, B. Salgado-Negret, J. Sardans, S. Shiodera, B. Shipley, A. Siefert, E. Sosinski, J.-F. Soussana, E. Swaine, N. Swenson, K. Thompson, P. Thornton, M. Waldram, E. Weiher, M. White, S. White, S. J. Wright, B. Yguel, S. Zaehle, A. E. Zanne, C. Wirth. 2011. TRY – a global database of plant traits. *Global Change Biology*, 17:2905–2935.

If data via TRY have made a decidedly minor contribution to the manuscript, offering authorship and citing original references in the main part of the manuscript may not be appropriate. However, these exceptions need to be discussed with the TRY database management in each case.

Apart from these general requirements of authorship, citation and acknowledgement, **please respect all individual agreements with data contributors made at the time of data requests!**